Proceedings of the 2010 3rd IEEE RAS & EMBS
International Conference on Biomedical Robotics and Biomechatronics,
The University of Tokyo, Tokyo, Japan, September 26-29, 2010

A Dynamic Neural Field Architecture for a Pro-active Assistant Robot

Manuel Pinheiro, Estela Bicho, Wolfram Erlhagen

Abstract— **We present a control architecture for non-verbal HRI that allows an assistant robot to have a pro-active and anticipatory behavior. The architecture implements the coordination of actions and goals among the human, that needs help, and the robot as a dynamic process that integrates contextual cues, shared task knowledge and predicted outcome of the human motor behavior. The robot control architecture is formalized by a coupled system of dynamic neural fields representing a distributed network of local but connected neural populations with specific functionalities. Different subpopulations encode task relevant information about action means, action goals and context in form of self-sustained activation patterns. These patterns are triggered by input from connected populations and evolve continuously in time under the influence of recurrent interactions. The dynamic control architecture is validated in an assistive task in which an anthropomorphic robot acts as a personal assistant of a person with motor impairments. We show that the context dependent mapping from action observation onto appropriate complementary actions allows the robot to cope with dynamically changing situations. This includes adaptation to different users and mutual compensation of physical limitations.**

I. INTRODUCTION

One of the current theoretical and experimental challenges in assistive robotics is how to develop autonomous robots able to assist people in a human-like way (reviews e.g.[13], [5], [21], [8], [26]). Humans prefer to interact with machines in the same way that they interact with other people[18]. This implies that in order to guarantee user acceptance, a personal assistant robot should be endowed with social and cognitive capacities that allow the human-robot interaction to be natural and efficient. In assistive tasks we continuously monitor the actions of the person with whom we interact, interpret them in terms of their underlying motor intentions, predict the concomitant outcome, and use these predictions to select an adequate action that complements and coordinates with the observed action [23], [16]. Imagine for instance the task of assisting a person during a meal. The way how a person grasps a certain object, e.g. a bottle of juice, allows the observer to infer the ultimate goal of the action. Depending on the grip type, the person may want to pour the juice on a glass or, alternatively, has the intention to hand it over. Being able to predict, at the time of grasping, the goal of the complete action allows the observer to timely grasp and hold out the glass, or to prepare to receive the bottle. The problem of action coordination may be solved using verbal communication but in some situations this kind of communication may slow down or even make the interaction impossible. Thus, non-verbal communication is considered essential. Reading the non-verbal cues inherent to every behavior favors the action transparency, the interaction robustness and bridge physical and cognitive (dis)abilities[6].

In this paper, we present results of our ongoing research to endow autonomous robots with cognitive capacities that will ultimately allow them to act as socially aware agents in assistive tasks to people with disabilities. Our strategy has been to develop an anthropomorphic robot that integrates in its control architecture recent experimental and theoretical findings about the neuro-cognitive mechanisms underlying perception and action in social contexts (e.g.[28], [12], [17], [10], for a review see [2], [23]).

It is believed that non-human primates and humans have within their neural structure, mechanism for mirroring observed actions [20], [28]. This mirror mechanism allows the observer to match an observed goal-directed motor act in his/her own motor repertoire, which the observer is familiar with. The system is called the mirror neuron system (MNS) and is thought to be the basic mechanism for action understanding and goal inference. Experiments taken with humans and monkeys revealed that the motor neurons that compose this mechanism have different degrees of specialization, and thus code different types of action related information. Some mirror neurons fire only when specific motor acts are observed, whereas others have a broader spectrum of activation. The mirror neurons that react to specific motor acts may be responsible for the representation of the means used to perform the action, like the type of grasp that it is used, while the mirror neurons that are less specific may code more abstract information about the action. It is thought that these neurons are responsible for action goal representation (for an overview see [19]). Moreover, the MNS is capable to represent actions within the cortical structures even when both agents have marked morphological differences [14]. What matters is the action effect (i.e. the underlying specific goal) an not so much the details of the movement trajectory that lead to this effect [15]. This result is very important because it allows us to apply the model of goal inference based on motor resonance in joint tasks that involve teammates with dissimilar embodiment like humans and robots. However, internally simulating the motor behavior of one's partner may not directly yield a

The present research was conducted in the context of the fp6-IST2 EU-project JAST (proj.nr. 003747) and partly financed by the FCT grant POCI/V.5/A0119/2005.

Manuel Pinheiro is with the Dept. of Industrial Electronics, University of Minho, Portugal manuelspinheiro@gmail.com

Estela Bicho (corresponding author) is with the Dept. of Industrial Electronics, University of Minho, Portugal estela.bicho@dei.uminho.pt

Wolfram Erlhagen is with the Dept of Mathematics and Applications, University of Minho, Portugal wolfram.erlhagen@mct.uminho.pt

978-1-4244-7707-4/10/$26.00 ©2010 IEEE

full understanding of the action goals. The same goal-direct action may have a different underlying goal depending on the context in which the action evolves. Thus, it is necessary to integrate additional contextual cues.

In our previous work we have developed and applied a dynamic field model of action understanding and complementary action selection that implements these ideas [4], [3]. It consists of a distributed network of local pools of neurons each with specific functionality. Self-stabilized activity patterns in these populations represent potential goals, context and potential action means to pursue the goals. Observed object-directed motor acts (e.g., grasping) together with contextual cues may trigger the propagation of activity through interconnected neural populations that constitute a learned chain of motor primitives directed towards a specific goal (e.g., reaching-grasping-placing at a particular position). The dynamic field model of action understanding and complementary action selection was tested in a joint construction task in which the human-robot team assembles a toy from its component parts knowing the construction plan. In this paper we validate the model in an assistive task scenario where the robot pro-actively helps a person, with motor disabilities, to drink. The focus of the results reported here is on dynamic action coordination among the robot and the human which have different motor limitations that have to be mutually compensated, flexible action selection in response to different contextual situations and different users (more active or more passive). Considering the human point of view, it may happen that non-verbal communication may not be enough to understand the robot's behavior. Thus, we enable the robot to reason aloud ir order to explain what is its understanding about what the human is trying to do, to give feedback to the human about what itself intends to do, to suggest how they can coordinate their actions and to communicate its own physical limitations.

The paper is organized as follows: Section II introduces the assistive task scenario and the robot platform. Section III gives an overview about the cognitive control architecture and presents the basic concepts of the dynamic field framework used to formalize and implement the architecture. The results of the human-robot interactions are described in section IV. The paper ends with a discussion of concepts and results and a short outlook.

II. ASSISTIVE TASK SCENARIO

One of the most important and elementary tasks performed by humans on their daily activities is eating and drinking. Here we consider an interaction scenario where the robot helps a human with physical limitations to drink (Fig. 1). The interaction scenario involves only two objects, a bottle and a glass, placed on a table and requires only a limited number of different motor actions to be performed by the human and the robot but is complex enough to show the impact of intention understanding on complementary action selection, and adaptation to different users. The table is divided in the middle by an imaginary line that defines the boundary between the human workspace (HWS) and

Fig. 1. The anthropomorphic robot acts as a pro-active personal assistant of a person, with motor impairments, that wishes to drink.

the robot workspace (RWS). The objects can be placed on the table with different arrangements and can have different states and orientations. More precisely, the bottle can be closed or open and the glass can be empty upright or inverted and full or empty. Based on the objects initial states and disposition on the table, the number and the nature of subtasks that must be accomplished to achieve the final goal of this joint action is different. Given this and based on the physical limitations of both agents, the interaction scenario has three main constraints: (1) it is assumed that the human user has a motor impairment that prevents him/her to perform the task alone; (2) the robot does not have enough dexterity in its hand to open the bottle, so the human is the only agent capable of removing the stopper/cup; (3) additionally and due to the human motor impairment, the human cannot open the bottle alone and needs the robot's help to remove the cup. The robot is physically unable to grasp such a small object, this means that to open the bottle the human and the robot are compelled to cooperate with each other. The cooperation between the teammates is also biased by the objects disposition on the table which may require handing over objects to one another. Both agents must cooperate with each other to compensate their mutual physical disabilities.
The robot was built in our lab [24]. It consists of a stationary torus on which a 7 DOFs AMTEC arm (Schunk GmbH) with a three finger dexterous gripper and a stereo camera head are mounted. A speech synthesizer (Microsoft Speech SDK 5.1) allows the robot to verbally communicate with the user. The information about object type, position and pose is provided by the camera vision system. The object recognition system combines color-based segmentation with template matching derived from earlier learning examples [27]. The same technique is also used for the classification of object-directed, static hand postures, such as reaching, grasping and grip type, and communicative gestures such as pointing or demanding an object. For the control of the arm-hand system we applied a global planning method in posture space that allows us to generate smooth and natural movements by integrating optimization principles obtained from experiments with humans [7].

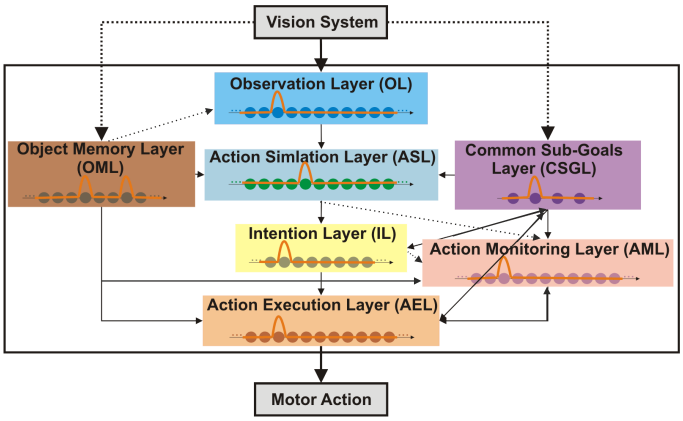

Fig. 2. The multi-layered cognitive control architecture for joint action. It implements a flexible context-dependent mapping between observed actions and executed actions.

III. COGNITIVE CONTROL ARCHITECTURE

Figure 2 presents schematically the robot cognitive control architecture which is inspired by known neuro-cognitive mechanisms underlying perception, reasoning and action in a social context (for details see [4], [3]).

The architecture can be seen as a network of interconnected pools of neurons, organized in different layers. Each layer is responsible for coding a different type of information and its dynamics enables the implementation of a dynamic process of action simulation, goal inference, action monitoring and complementary action selections[4]. The control architecture receives input information from the vision system. The information provided by the vision system produces activation patterns in specialized pools of neurons within the Observation Layer (OL), where the motor acts observed by the robot are translated into motor primitives. In the Object Memory Layer (OML) specific pools of neurons represent the objects of interest in terms their general position within the agents' workplaces. The Common Sub-Goals Layer (CSGL) consists of three sub-layers and endows the robot with an internal dynamic representation that codes which sub-goals have already been executed (past sub-goals), which sub-goals can currently be performed (present sub-goals) and future sub-goals.

Once the patterns of activation on the input layers become self-sustainable, they elicit an activation pattern in the Action Simulation Layer (ASL), where the robot internally simulates the action performed by its partner by activating the goal directed action that corresponds to the sequence of observed motor acts. The patterns of activation in ASL and CSGL activate the subpopulation in the Intention Layer (IL) representing the current intention of the teammate. The intention along with the information provided by OML and CSGL provide input to the Action Execution Layer (AEL), which through competition selects the most appropriate complementary robot's behavior. The Action Monitoring Layer provides also input to AEL and it is also a key factor in the process of selecting the most proper complementary action. This layer represents mismatches between the inferred intention of the human partner and the subgoals that are currently possible[4].

A. The Dynamic Neural Field framework

Each layer is implemented using the theoretical framework of the Dynamic Field Theory [25], [9], [22]. In each layer ($i = ASL, AEL, ...,$), the activity $u_i(x,t)$ at time t of a neuron at field location x is described by the following integro-differential equation ([1]):

$$\tau_i \frac{\delta u_i(x,t)}{\delta t} = -u_i(x,t) + S_i(x,t) + \int w_i(x-x') f_i(u_i(x',t)) dx' + h_i \quad (1)$$

This equation is the mathematical model for a one-dimensional field of lateral-inhibition type. The neural fields of lateral-inhibition type are homogeneous fields that include both excitatory and inhibitory neurons. Within these fields, the temporal dynamics of individual neurons is neglected over the overall behavior of the entire population. Parameters τ_i and $h_i < 0$ define the time scale and the resting level of the dynamic field, respectively, and $S_i(x,t)$ represents the external input applied to the field at location x and time t. $u_i(x,t)$ represents the activity of a neuron coding the field location x. The integral term implements the convolution of the function containing the weights of the internal connections between neurons and the non-linear output function $f(u)$, that only allows neurons that are active above a threshold to contribute to that same internal interactions. Since the field is of lateral-inhibition type, by convention the excitatory connections dominate at proximal distances and the inhibitory connections dominate at greater distances [1]. This means that the excitatory or inhibitory interaction between two neurons coding field location x and x' respectively, only depends on their distance, $w(x-x')$. The interaction behavior of all neurons can thus be modeled by a Gaussian function minus a constant value, w_{inhib}:

$$w(x-x') = Ae^{-\frac{(x-x')^2}{2\sigma^2}} - w_{inhib} \quad (2)$$

where $A > 0$ and $\sigma > 0$ define, respectively, the amplitude and standard deviation, and the constant $w_{inhib} > 0$ represents the global inhibition that the active neurons carry on the rest of the field. Only the neurons that received enough amounts of input to become positively active are able to transmit information to the down-stream systems and to contribute to the internal interactions of the neural field. Additionally, it is assumed that when the neurons pass over the activation threshold, they all fire at their maximum rate, regardless of the magnitude of the input pattern. To model these properties the field dynamics must become highly non-linear, which is achieved through the sigmoid function showed in equation 3:

$$f(u) = \frac{1}{1+e^{-\beta(u-u_0)}} \quad (3)$$

where u_0 is the threshold and $\beta > 0$ is the slope parameter [9].

The summed input from connected fields u_l is given as $S_i(x,t) = k\sum_l S_l(x,t)$. The parameter k scales the total input to a certain population relative to the threshold for triggering a self-sustained pattern. This guarantees that the inter-field couplings are weak compared to the the recurrent interactions that dominate the field dynamics (for details see [9]). The scaling also ensures that missing or delayed input from one or more connected populations will lead to a subthreshold activity distribution only. The input from each connected field u_l is modeled by Gaussian functions

$$S_l(x,t) = \sum_m \sum_j a_{mj} c_l(t) \exp(-(x-x_m)^2/2\sigma^2) \qquad (4)$$

where $c_l(t)$ is a function that signals the presence or absence of a self-stabilized activation peak in u_l, and a_{mj} is the inter-field synaptic connection between subpopulation j in u_l to subpopulation m in u_i. Inputs from the vision system are also modeled as Gaussians for simplicity.

IV. RESULTS

In the following we validate the dynamic control architecture by presenting results of human-robot interactions in the assistive task. The examples represent different video snapshots that are chosen to illustrate the impact of action observation on complementary action selection from the perspective of the robot. The videos can be found at http://dei-s1.dei.uminho.pt/pessoas/estela/BioRob2010.htm. We will try to briefly illustrate how the overall system works and how the actions performed by the human, the context in which the action is executed and even the 'personality' of the human can influence the robot's decision making process. It is important to stress that in these tests the following assumptions are made: *i)* it is assumed that the human wants to drink but he suffers from a motor impairment that prevents him to perform the task alone; *ii)* the robot has prior knowledge about the task, i.e. the robot knows that first it is necessary to open the bottle and/or turn the glass in upright position before filling it; *iii)* the robot knows that observing a grasping of the bottle or glass from above (above grip, AG) means that the human most likely is going to handover the object; *iv)* grasping the bottle from the side (side grip, SG) means that probably the human will try to pour the juice in the glass; *v)* grasping the glass from the side (side grip) means that the human will likely try to invert it or to drink depending on its state. It is important to strength that there is not a one to one mapping. In previous work we have shown that an imitation learning paradigm can be used to transfer the knowledge about this specific grip-goal relation from a human teacher to the robot that takes into account the context[11].

A. Impact of Action simulation, Goal Inference and Action Monitoring on Action Execution

The robot is capable of acquiring the object positions and states, and the motor primitives used by the human to interact with the environment, through the vision system. That information produces activation patterns on specialized

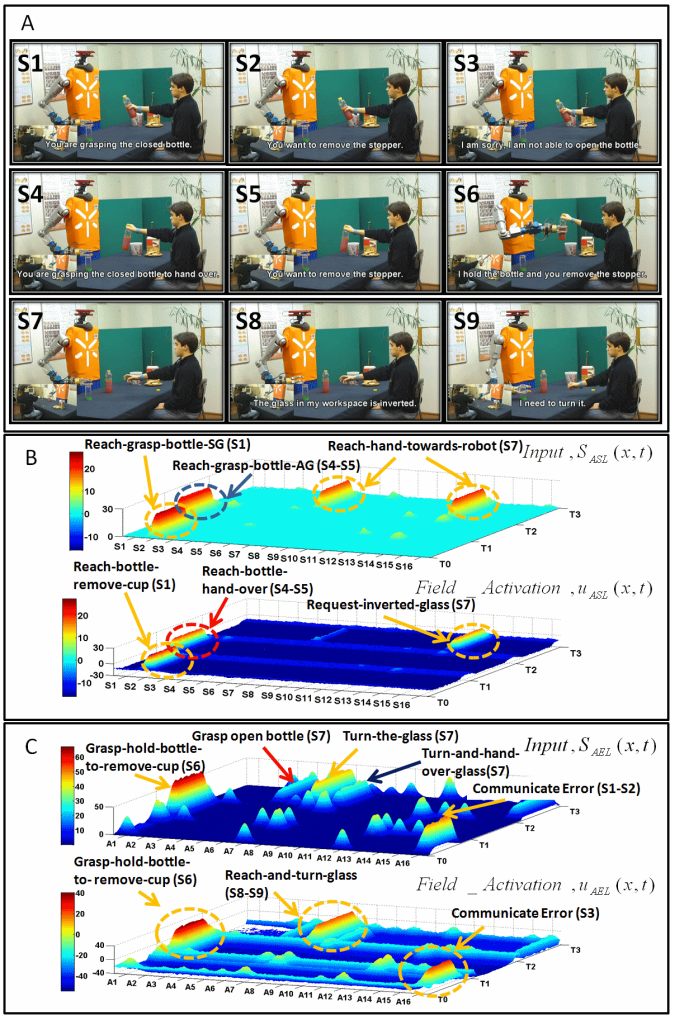

Fig. 3. An example that shows the impact of goal inference and action monitoring on the process of complementary action selection. Panel A: video snapshots. Panel B: Temporal evolution of the input (top) to ASL and activity in ASL (bottom). Panel C: Temporal evolution of the input to AEL (top) and activity in AEL (bottom).

pools of neurons in the OL, in the OML and in the CSGL. The motor and the context information, coded by the OL, the OML and the CSGL layers, enable the robot to simulate the observed action at the ASL level and then obtain its underlying intention. Thus, the decision cycle is always composed by the action simulation and intention inference processes. These two processes allow the robot to continuously track and relate actions with their underlying context, monitoring their outcomes. Figure 3 shows the impact of action simulation, goal inference and action monitoring on action execution, which may include overt motor behavior and/or speech. The robot monitors the behavior of the human agent, selecting the most appropriate action, taking always into account the interaction scenario constraints and the agents' physical limitations.

Due to his supposed motor disability, the human is not capable of opening the bottle on his own, so he grasps and holds out out the bottle (Panel A, snapshot *S1*) as a

request for the robot's help. The action performed by the human is decomposed into its motor primitives, i.e., `Reach`, `Grasp` with `SG` (Side Grip), object `Closed-Bottle`. This information (OL) along with the contextual information provided by the OML and the CSGL activates de pool of neurons coding the goal direct action-chain $Reach-Grasp-SG-Closed-Bottle$ in ASL layer (panel B $(T0-T1)$). After that, the robot infers the intention of removing the stopper, which is represented by an activation peak in the IL. However, the robot is not able to grasp and rotate such a small object and the control architecture produces an error in the Action Monitoring Layer (AML). The pattern of field activation within AML, coding that error, directly influences the action selection process in the AEL. As a result of that, the robot verbally communicates to the human its own physical limitation (panel A, snapshots $S2$ and $S3$). As it can be seen in panel C $(T0-T1)$, initially there are some input stimuli at different positions in the AEL, however all of them disappear and $Communicate-Error$ neuronal population wins. The $Communicate-Error$ activation enables the reproduction of the error audio messages. The audio message that is displayed depends directly on the error that is detected by the AML, i.e., depends on the winning activation within the AML due to its internal competition.

B. Dynamic Action Coordination

The two agents have physical limitations that need to be mutually compensated. For instance, the robot cannot remove the stopper/cup and the human can but only when helped by the robot, i.e., to remove the stopper the robot has to hold the bottle while the human removes it. Figure 3, snapshots $S4$ to $S6$ in panel A, show one example of the process of action coordination between the two agents that allow coping with their physical limitations. The human reaches and grasps the closed bottle (CB) from above (panel A, snapshots $S4$ and $S5$). As previously, the motor action is decomposed in its elementary motor primitives, i.e. `Reach`, `Grasp` with `Above Grip` (AG) object `close bottle`. This sequence of motor primitives is associated with the goal directed action of handing over the CB in the ASL, which in turn is also closely related with the intention of removing the stopper in IL. Contrary to the previous situation (snapshots $S1-S3$), the action of grasping the bottle from above enables the robot to safely grasp and hold it out so the human can remove the stopper. Given this, the robot decides to hold the bottle for the human and gives him instructions that he may remove the stopper (panel A, snapshot $S6$).

Attending to the human desires does not come from a purely reactive behavior but rather comes from some level of reasoning about the interaction scenario. Snapshots $S7$ to $S9$ of figure 3 provide a good example and reveal once again the importance of the contextual data on the final decision (in this particular case the effect of the OML in the AEL). The human holds out his empty hand towards the robot which infers that the human is requesting the (inverted) glass (IG) in its workspace. In this situation two main goal directed actions actions compete in AEL for expression in overt behavior, namely, $Reach-and-turn-glass$ $(A6)$ and $Reach-Turn-and-hand-over-glass$ $(A7)$. The selection of the $A6$ instead of $A7$ comes from the fact that within the OML it is encoded an open bottle (OB) in the robot's workspace (RWS). Based on this information, the action of handing over the glass is inhibited and the evolution of the activity in AEL is biased to select the action of reaching the inverted glass (IG) and place it once more in the RWS. This situation evidences that in each action selection process there is always a basic reasoning mechanism. In this example, since the robot has the bottle in its workspace, and it is opened, it is more efficient to hand over the glass but only when filled by the robot itself.

C. Fluency in the Interaction

Like human, robots must able to perceive and predict the intentions underlying ongoing actions, shaping its behavior. The capability to predict the human intention is important for fluent and efficient interaction, and is a fundamental feature to turn the robot into an effective socially aware assistant robot.

Figure 4 presents two different situations in which the flexibility of the decision process is crucial to ensure consistency and fluidity in the interaction. At snapshot S10 (in panel A) the robot has the Open Bottle (OB) and the Empty Glass (EG) in its workspace, this means that there are already two sub-goals that have been fulfilled (i.e. $Open-Bottle$ and

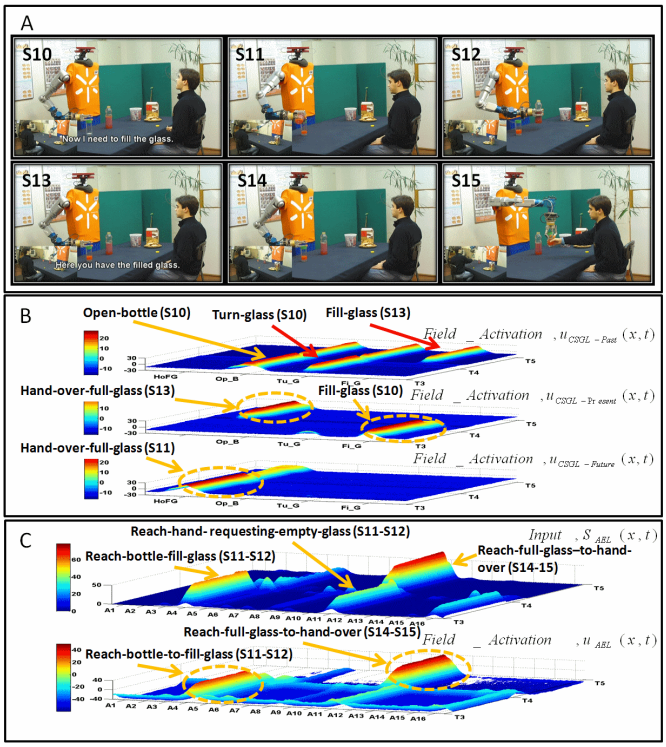

Fig. 4. The impact of CSGL in anticipatory action selection. Panel A: continuation of the snapshots of video in Fig. 3. Panel B: Field activity in CSGL Past, Present and Future layers, respectively. Panel C:Temporal evolution of the input to AEL (top) and activity in AEL (bottom).

Turn – Glass) and other two sub-goals (i.e. *Fill – glass* and *Hand – over – full – glass*) that need to be accomplished.

In panel B (figure 4) one can observe which sub-goals have been met initially, they are represented at the *GSGL – Past* (*Open – bottle* and *Turn – glass* sub-goals), and the sub-goals that should be under the attention of the robot arise at the *GSGL – Present* and *GSGL – Future* (i.e. *Fill – glass* and *Hand – over – full – glass*, respectively). Since it is necessary to satisfy first the *Fill – glass* sub-goal over the *Hand – over – full – glass* sub-goal, an activation peak emerges in the *GSGL – Present* coding that priority. Due to its current low priority, the *Hand – over – full – glass* sub-goal appears at the *GSGL – Future*, figure 4 panel B ($T3-T4$). The activation patterns in CSGL benefit all actions in AEL that satisfy the goal of filling the glass. The action selection process is thus biased to produce an output that drives the robot to *Reach – grasp – open – bottle – fill – glass* (figure 4, snapshot $S11$ and $S12$).

At snapshot $S13$ the robot faces a similar situation and it has to decide on its own which path the interaction should take. From the panel B ($T4-T5$) of figure 4 one can observe that now the *Fill – glass* sub-goal is at the *GSGL – Past* and the *Hand – over – full – glass* is now at the *GSGL – Present*. As a consequence of that, the robot's behavior is different from the previously observed. It now reaches the (full) glass and hands it over to the human (figure 4 snapshot $S14$ and $S15$ and *Reach – full – glass – to – hand – over* from panel C ($T4-T5$)). These two situations show that the robot is capable of flexibly decide what action must be performed and that shared task knowledge (CSGL) plays a very important role in that flexibility.

D. Adaptation to different users

In order to study the capability of the robot to take the initiative when interacting with a human, some experiments where carried out where the human had a passive attitude. The interaction scenario is composed by a closed bottle (CB) and an empty glass (EG) in the upright position. The open bottle and the glass are located in the robot's and user's workspace respectively. The sub-goal of turning the glass is, at the outset of the interaction, accomplished (see figure 5, panel B ($T0-T1$) *CSGL – Past*).

From the *CSGL – Present* the need of opening the bottle is extracted and the decision in AEL is biased toward the robot asking for help to open the bottle. As can be seen in panel C, time interval $T0-T1$, in the AEL emerges the action of *Grasp – hold – bottle – to – remove – cup* (see also snapshots $S1$ to $S3$ in panel A, figure 5). After that, the *Open – bottle* sub-goal disappears from the *CSGL – Present* and emerges at the *CSGL – Past* and the *Fill – glass* previously at the *GSGL – Future* rises at the *CSGL – Present*, setting the next interaction priority (panel B ($T1-T2$) *CSGL – Present* and *CSGL – Future*). Now, the robot has the open bottle (OB) in its workspace while the empty glass (EG) remains at the human's workspace. In this situation the robot has two possible actions competing for expression in overt behavior, 'handover the bottle' to the human or 'request

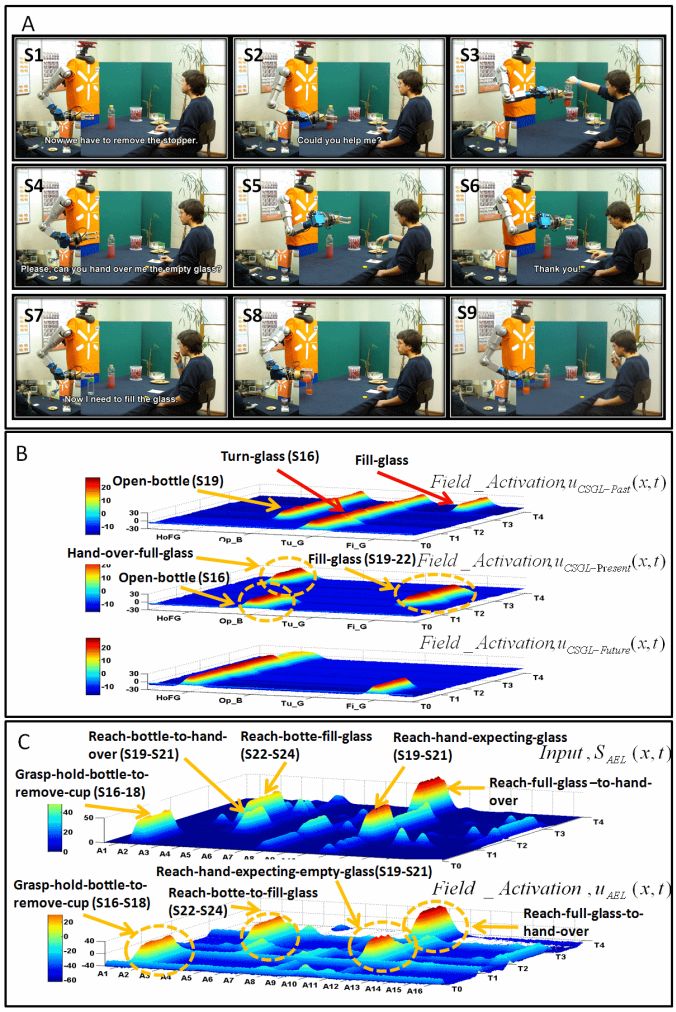

Fig. 5. An example that illustrates the capability of the robot to take the initiative when interacting with a passive person. Panel A: Video snapshots. Panel B: Field activity in CSGL Past, Present and Furure layers, respectively. Panel C: Temporal evolution of the input to AEL (top) and activity in AEL (bottom).

the glass' (panel C, ($T1-T2$)). It decides to request the glass and holds out its hand towards the human (panel A, snapshots $S4$ to $S6$). As can be observed, with this action the goal of filling the glass has not yet been satisfied, and this task sub-goal remains active from the current interaction to the next one (panel B,($T2-T3$) *CSGL – Present*). At the time between $T2$ and $T3$ the robot fills the empty glass (panel A, snapshots $S7$ to $S9$), the *Fill – glass* sub-goal disappears from the *CSGL – Present* and rises at the *CSGL – Past* and then the *Hand – over – full – glass* is set as the present sub-goals (panel B ($T3-T4$) *CSGL – Present*), which in turn triggers in the AEL the goal directed action of *reach – grasp – full – glass – to – handover*.

V. Discussion and Conclusion

In order to interact with humans in a social context and particularly in assistive tasks to people with disabilities, the robot must be capable of extracting meaning from what it

observes. In the assistive task here used to test the robot capabilities, the robot has to extract information from the interaction scenario (i.e. obtain the objects positions and states) and also from its partner by tracking his goal directed actions, the context in which they occur, and infer the underlying motor intentions. Only based on these can the robot select an adequate complementary actions that better facilitates the interaction and serves the human user.

The capability of human and non-human primates to understand actions performed by other individuals is thought to be closely linked with their motor repertoire. It is thought that primates extract meaning from actions based on their one motor experience, i.e., from the actions that are somehow coded in their cortical structures. The MNS seems to be particularly important in this issue because it is thought to be the fundamental mechanism for goal-directed actions representation within the cortex, and thus for interaction itself. The cognitive architecture for human-robot interaction here presented tries to implement some of the elementary features of the MNS in an attempt to endow the robot with rudiments of action understanding and goal inferencing. When the human performs a goal-directed motor act, that action can be characterized by the way that the human interacts with the environment (objects) and by its target (objects or the robot itself). It is this information that enables the robot to first represent actions in specialized pools of neurons at the OL and after that 'simulate' them (through activation of the corresponding goal directed motor-chains) at the ASL level.

The robot is capable of acquiring the object positions and states, and the motor primitives used by the human to interact with the environment, through the vision system. This information produces activation patterns on specialized pools of neurons in the OL, in the OML and in the CSGL. The motor primitives in the OL normally have several actions attached to them, so in order to simulate correctly what action was really performed, it is necessary to combine that data with contextual information and task knowledge, which are coded in OML and CSGL, respectively. By having a representation of the action and its grounding context in its cognitive structure, the robot is also capable of extracting the underlying intentions of actions (at the IL). The ability to simulate actions, infer action underlying intentions and accommodate the environmental information in the decision making process enables the robot to understand the human and, more importantly, predict the outcomes of those actions. Clear evidences of this are the action monitoring and flexible action allocation abilities demonstrated by the robot. Additionally, the ability to predict the action consequences allowed the robot a behavior far more complex than a purely reactive agent. In some situations the robot did not satisfy the human's desires, that did not imply a wrong behavior by itself, but instead a step forward in the interaction, as illustrated in figure 3 snapshots $S7$ to $S9$.

Moreover, the role of the OML and the CSGL becomes particularly important when the human for some reason becomes/is more passive. The information coded in these layers enabled the robot to engage in interaction with the human even when he was more passive or totally passive. This feature is of particular importance if one considers the possible application of robots as socially assistive agents, in healthcare facilities or in home environments, providing care to people with motor or cognitive disabilities.

Another important feature is the fact that the robot acquires all the necessary information from the human without using any verbal communication. The verbal communication in scenarios where a fast interaction is needed may be a disadvantage or even inconceivable, for instance when dealing with people with serious cognitive disabilities. The use the non-verbal communication provides a common ground of understanding between teammates and enables the robot to interact with people from different cultural background and geographies without changing anything. The verbal communication exists only from the robot to the human in order to provide an outward feedback of its internal state and to help the human to understand what the robot is doing. The robot's verbal expression can be seen as an embodied dialog, i.e., it comes only from what the robot observes and does. Within the cognitive architecture every pool of neurons of the ASL, the IL, the AML and the AEL has a single verbal expression associated a priori to it. The dialogue is constructed based on what the robot simulates in the ASL, what if infers in the IL and AML and on what it decides in the AEL, without any external interference.

At first sight it may look like that the capabilities of the robot presented could have been implemented by an intricate state machine. One very important difference is the aspect of time, i.e. the timing at which the decisions and actions of the human evolve play a role in the robot's interpretation and decision processes. In the dynamic neural field architecture the decision process linked to complementary actions unfolds over time under multiple influences which are themselves modelled as dynamic representations with proper time scales. This is the basis of flexible behaviour in dynamic joint action conditions. As was shown in the results section, the absence or delay of information about for instance the co-actor's motor intention will automatically lead to a decision that does not take into account the co-actor.

An obvious extension of the present work is to test the robot in more complex assistive tasks, and to endow the assistant robot with learning capabilities that will allow it to learn from, and adapt to users that for example use different means (i.e. different motor acts) when performing the same task, and thus become a more adequate social partner. In previous work we have made the first steps to show that by using correlation based learning rules with a gating that signals the success of behavior it is possible to evolve mirror-like representations that support an action understanding capacity [11], [10].

VI. ACKNOWLEDGMENTS

The authors gratefully acknowledge the help from Eliana Silva, Emanuel Sousa, Flora Ferreira, João Ferreira, Joaquim

Silva, Luís Louro, Nzoji Hipolito, Rui Silva, Tiago Malheiro and Toni Machado, and the reviewers' comments.

REFERENCES

[1] S. Amari, "Dynamics of pattern formation in lateral-inhibitory type neural fields," *Biological Cybernetics*, vol. 27, pp. 77–87, 1977.

[2] H. Bekkering, E. de Bruin, R. Cuipers, R. D. Newman-Norlund, H. T. Van Schie, and R. G. J. Meulenbroek, "Joint action: Neurocognitive mechanisms supporting human interaction," *Topics in Cognitive Science*, vol. 1, pp. 340–352, 2009.

[3] E. Bicho, W. Erlhagen, L. Louro, and E. Costa e Silva, "Neuro-cognitive mechanisms of decision making in joint action: a human-robot interaction study," *submitted to Human Movement Science (minor revisions requested)*, 2010.

[4] E. Bicho, L. Louro, N. Hipolito, and W. Erlhagen, "A dynamic field approach to goal inference and error monitoring for human-robot interaction," in *Proceedings of the 2009 International Symposium on New Frontiers in Human-Robot Interaction*, K. Dautenhahn, Ed. AISB 2009 Convention, Heriot-Watt University Edinburgh, 2009, pp. 31–37.

[5] C. Breazeal, "Social interactions in hri: The robot view," *IEEE Transactions on Systems, Man and Cybernetics- Part C: Applications and Reviews*, vol. 34, no. 2, pp. 181–186, 2004.

[6] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, and M. Berlin, "Effects of nonverbal communication on efficiency and robustness in human-robot teamwork," in *Proceedings of IEEE/RJS. Int. Conference on Intelligent Robots and Systems (IROS'2005)*, 2005, pp. pp.383–388.

[7] E. Costa e Silva, E. Bicho, W. Erlhagen, and R. Meulenbroek, "Human-like collision-free arm movements for human-robot collaboration," *Submitted.*, 2010.

[8] K. Dautenhahn, "Socially intelligent robots: dimensions of human-robot interaction," *Phil. Trans. R. Soc. B*, vol. 362, pp. 679–704, 2007.

[9] W. Erlhagen and E. Bicho, "The dynamic neural field approach to cognitive robotics," *Journal of Neural Engineering*, vol. 3, pp. R36–R54, 2006.

[10] W. Erlhagen, A. Mukovskiy, and E. Bicho, "A dynamic model for action understanding and goal-directed imitation," *Brain Research*, vol. 1083, pp. 174–188, 2006.

[11] W. Erlhagen, A. Mukovskiy, E. Bicho, G. Panin, C. Kiss, A. Knoll, H. van Schie, and H. Bekkering, "Goal-directed imitation for robots: a bio-inspired approach to action understanding and skill learning," *Robotics and Autonomous Systems*, vol. 54, pp. 353–360, 2006.

[12] L. Fogassi, P. F. Ferrari, B. Gesierich, S. Rozzi, F. Chersi, and G. Rizzolatti, "Parietal lobe: from action organization to intention understanding," *Science*, vol. 308, pp. 662–667, 2005.

[13] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," *Robotics and Autonomous Systems*, vol. 42, pp. 143–166, 2003.

[14] V. Gazzola, G. Rizzolatti, B. Wicker, and C. Keysers, "The anthropomorphic brain: The mirror neuron system responds to human and robotic actions," *Neuroimage*, vol. 35, no. 4, pp. 1674–1684, 2007.

[15] M. Iacoboni, I. Molnar-Szakacs, V. Gallese, G. Buccino, J. Mazziotta, and G. Rizzolatti, "Grasping the intentions of others with ones own mirror neuron system," *PLoS Biol*, vol. 3, no. 3, p. e79, 2005.

[16] R. D. Newman-Norlund, M. L. Noordzij, R. G. J. Meulenbroek, and H. Bekkering, "Exploring the basis of joint action: Coordination of actions, goals and intentions," *Social Neuroscience*, vol. 2, pp. 48–65, 2007.

[17] R. D. Newman-Norlund, H. T. van Schie, A. M. J. van Zuijlen, and H. Bekkering, "The mirror neuron system is more active during complementary compared with imitative action," *Nature Neuroscience*, vol. 10, pp. 817–818, 2007.

[18] B. Reeves and C. Nass, *The media equation- how people treat computers, television and new media like people and places*. Cambridge, UK: Cambridge University Press, 1996.

[19] G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annual Review of Neuroscience*, vol. 27, pp. 169–192, 2004.

[20] G. Rizzolatti, L. Fogassi, and V. Gallese, "Neurophysiological mechanisms underlying the understanding and imitation of action," *Nature Reviews*, vol. 2, pp. 661–670, 2001.

[21] S. Schaal, "The new robotics: towards human-centered machines," *HFSP Journal*, vol. 1, pp. 115–126, 2007.

[22] G. Schöner, "Dynamical systems approaches to cognition," in *The Cambridge Handbook of Computational Psychology*, R. Sun, Ed. Cambridge University Press, 2008, pp. 101–125.

[23] N. Sebanz, H. Bekkering, and G. Knoblich, "Joint action: bodies and minds moving together," *Trends in Cognitive Sciences*, vol. 10, pp. 70–76, 2006.

[24] R. Silva, E. Bicho, and W. Erlhagen, "Aros: An anthropomorphic robot for human-robot interaction and coordination studies," in *Proceedings of the CONTROLO2008 - 8th Portuguese Conference on Automatic Control*, 2008, pp. 819–826.

[25] J. Spencer and G. Schoner, "Bridging the representational gap in the dynamic systems approach to development," *Developmental Science*, vol. 6, no. 4, pp. 392–412, 2003.

[26] A. Tapus, M. Mataric, and B. Scasselatti, "The grand challenges in socially assistive robotics," *IEEE Robotics and Automation Magazine*, vol. 14, no. 1, pp. 35–42, 2007.

[27] G. Westphal, C. von der Malsburg, and R. P. Würtz, "Feature-driven emergence of model graphs for object recognition and categorization," in *Applied Pattern Recognition, Sudies in Computational Intelligence Vol. 91*, B. H, A. Kandel, and M. Last, Eds. Springer Verlag, 2008, pp. 155–199.

[28] M. Wilson and G. Knoblich, "The case for motor involvment in perceiving conspecifics," *Psychological Bulletin*, vol. 131, pp. 460–473, 2005.